

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平7-181888

(13) 公開日 平成7年(1995)7月21日

(51) Int. CL <sup>4</sup>	識別記号	片内整理番号	P I	技術表示箇所
G 0 9 B 21/04				
A 6 1 F 5/53		7108-4C		
G 0 8 P 15/18	5 6 0 G	9071-5L		
G 1 0 L 3/00		E		

審査請求 未請求 請求項の数 3 O L (全 10 頁)

(21) 出願番号	特願平5-327827	(71) 出願人	000004228 日本電信電話株式会社 東京都千代田区内幸町一丁目1番6号
(22) 出願日	平成5年(1993)12月24日	(72) 発明者	平岩 明 東京都千代田区内幸町1丁目1番6号 日 本電信電話株式会社内
		(72) 発明者	曾根原 登 東京都千代田区内幸町1丁目1番6号 日 本電信電話株式会社内
		(74) 代理人	弁護士 伊東 孝彦

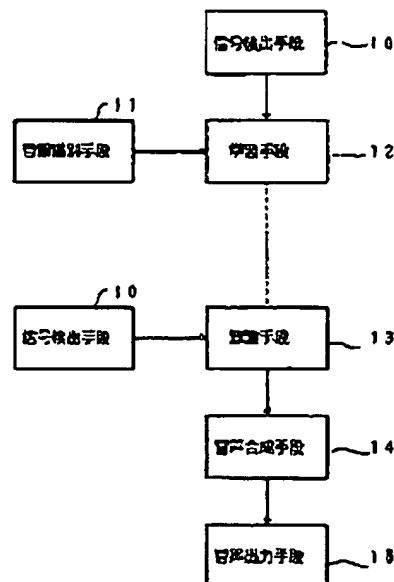
(54) 発明の名称 発声代行装置

(57) 要約

【目的】 本発明の目的は、何等かの理由で正常に発声できない人のために、口周辺の筋電位に基づいた合成音声を出力する発声代行装置を提供することである。

【構成】 本発明は、信号検出手段10により検出された信号を入力パターンとし、音声識別手段11により生成された音素に対応する識別信号を教師パターンとして学習を行う学習手段12と、信号検出手段10により検出された信号を入力パターンとし、対応する発声音素信号を認識して出力する認識手段13と、発声音素信号を合成して音声に変換する音声合成手段14と、音声信号を音声として出力する音声出力手段15とを有する。

本発明の構成図



(2)

特開平7-181888

1

## 【特許請求の範囲】

【請求項1】 発声動作による筋肉の動きにより発生する信号を検出する信号検出手段と、

該発声動作を行うと共に所望の音節に対応する識別信号を生成する音節識別手段と、

該信号検出手段により検出された信号を入力パターンとし、該音声識別手段により生成された該音節に対応する識別信号を教師パターンとして学習を行う学習手段と、該信号検出手段により検出された信号を入力パターンとし、該入力パターンに対応する発声音節信号を認識して出力する認識手段と、

該認識手段により出力された該発声音節信号を合成して音声に変換する音声合成手段と、

該音声合成手段により変換された該音声信号を音声として出力する音声出力手段とを有することを特徴とする発声代行装置。

【請求項2】 前記信号検出手段は

皮膚の動きにより信号を検出するための複数の皮膚表面電極と、

該皮膚表面電極から検出された筋電信号を増幅するための増幅手段と、

該増幅手段により増幅された信号の低周波成分と高周波成分を除去するフィルタ手段と、

該フィルタ手段により濾波された信号をパワースペクトルに変換する信号変換手段とを含む請求項1記載の発声代行装置。

【請求項3】 前記学習手段及び認識手段は、ニューラルネットワークを用いる請求項1記載の発声代行装置。

【発明の詳細な説明】

【0001】

【産業上の利用分野】本発明は、発声代行装置に係り、特に なんらかの病気で喉頭の切除を受け発声ができない無喉頭者に対して、音節を発声したいと意図した場合に頬部及び下顎部に装着された皮膚表面電極から検出された皮膚表面筋電パターンを認識して、音声を図示した音節を認識し、認識した音声を生声合成装置により人工的に発声する筋電パターンを認識して音声合成を行う人工咽頭に変わる発声代行装置に関する。

【0002】

【従来の技術】なんらかの病気で喉頭を失った無喉頭者は、正常な発声ができなくなるが、これらの人の発声を補助するのに、喉の外に振動子を当てて喉頭に音源を送り込む方法や、パイプを口の中に挿入して直接送る方法がある。

【0003】また、発声者の唇の動きを画像により取り込み、認識処理するコンピュータによるリップリーディングがある。

【0004】さらに、口の周囲の筋内の動きに伴って発生する筋電信号を処理して母音の種類を識別するという研究がある。これは、文献「Noboru Sumi et. al., A

2

speech Employing a Speech Synthesizer Vowel Discrimination from Perioral Muscles Activities and Vowel Production, IEEE transactions on Biomedical Engineering, Vol. 32, No. 7, pp485-490」に示されており、筋電信号をバンドパスフィルタを通した後、関節の交差回数をカウントして5母音(a, i, u, e, o)を弁別するものである。

【0005】

【発明が解決しようとする課題】しかしながら、上記従来の振動子を用いて音源を送り込む方法や、パイプを用いる方法は、いずれも音が不自然でブザーのような音になってしまうため、他の者がその音を音声としては認識しにくい。

【0006】また、コンピュータによるリップリーディングは、画像処理による認識率の低いことや、コンピュータの認識能力に依存する。

【0007】さらに、筋電信号により母音を識別する方法は、母音は識別できるが、子音については完全に識別することができないという問題がある。

【0008】本発明は、上記の点に鑑みなされたもので、上記従来の問題点を解決し、何等かの理由で、正真正正に発声できないもののために、口周辺の筋電位に基づいた合成音声を生声出力することができる発声代行装置を提供することを目的とする。

【0009】

【課題を解決するための手段】図1は、本発明の原理構成図である。

【0010】発声動作による筋肉の動きにより発生する信号を検出する信号検出手段10と、発声動作を行うと共に所望の音節に対応する識別信号を生成する音節識別手段11と、信号検出手段10により検出された信号を入力パターンとし、音節識別手段11により生成された音節に対応する識別信号を教師パターンとして学習を行う学習手段12と、信号検出手段10により検出された信号を入力パターンとし、入力パターンに対応する発声音節信号を認識して出力する認識手段13と、認識手段13により出力された発声音節信号を合成して音声に変換する音声合成手段14と、音声合成手段14により変換された音声信号を音声として出力する音声出力手段15とを有する。

【0011】また、本発明の信号検出手段10は、皮膚の動きにより信号を検出するための複数の皮膚表面電極と、皮膚表面電極から検出された筋電信号を増幅するための増幅手段と、増幅手段により増幅された信号の低周波成分と高周波成分を除去するフィルタ手段と、フィルタ手段により濾波された信号をパワースペクトルに変換する信号変換手段とを含む。

【0012】また、本発明の学習手段12及び認識手段13は、ニューラルネットワークを用いる。

【0013】

50

(3)

特開平7-181888

3

【作用】本発明は、認識部に学習性と雑音に強い認識率の高い処理機構であるニューラルネットワークを導入し、このニューラルネットワークに学習データとなる発声動作から得られる信号と、発声動作に対応する音節対応信号を教師データとして入力して学習処理を行う。学習後、ユーザの発声動作により、複数の皮膚表面電極から検出された筋電信号をFFT（高速フーリエ変換）により周波数スペクトルに変換し、変換されたスペクトルを認識して、認識された音声音節を認識し、ユーザの発声動作による筋電信号により音声合成を行い、音声出力する発声代行装置を提供し、無咽喉者が日常生活で発声動作を行うことにより自然な音声に近い合成音声を出力することができる。

【0014】

【実施例】以下、本発明の実施例を図面と共に説明する。

【0015】最初に本発明の概要を説明する。

【0016】図2は、本発明の概要を説明するための図である。

【0017】まず、同図（A）に示すように、ユーザは発話装置を用いて所望の音節を口を動かすことにより発声動作を行い、その動作に応じた信号が検出される。この信号を学習データとするために信号変換処理を行い、ニューラルネットワークに入力する。さらに、キーボード等の音節に対応する信号を作成できる機器を用いて発声動作に伴って識別信号を生成し、これを学習処理の教師データとする。ニューラルネットワークは入力された学習データと教師データにより学習を行う。

【0018】学習処理を終了し、学習処理を行ったユーザが発話装置より発声動作を行い、この動作により得られた信号をニューラルネットワークに入力できるように変換処理を行い、認識処理を行うためにニューラルネットワークに入力する。ニューラルネットワークは、先に学習しているので、入力された信号の認識を行い、認識結果を出力する。認識結果である信号を音声合成し、音声として出力する。

【0019】本実施例においては、発話装置を用いて発声動作を行うが、これは、ユーザが音声を出すのではなく、口の周辺の筋肉を動かすことにより発声する信号を検出し、この信号に基づいて、学習処理及び認識処理を行うものである。

【0020】図3は、本発明の一実施例のシステムを説明するための図である。

【0021】同図に示すように、発話装置1は、キーボード2と接続され、利用者は、発話装置1を片手で持ち、他方の手でキーボード2より、発声した音声についての指示入力を行う。発話装置は、同図に示すように非絶縁性のプラスチック等で構成された筐体であり、発話者の口付近と咽の部分に接触するように曲線からなる形状に構成されている。

4

【0022】図4は、本発明の一実施例の発話装置を裏面から見た図である。

【0023】同図は、図2に示す発話装置1を裏面から見た場合の図であり、発話装置1の裏面には、皮膚表面電極31～42が配設され、各電極31～42は、電極31と32、電極33と34というように双極誘導するために対になっており、発話者の顔の表面に装着することにより筋電信号を得る。

【0024】発話者は、上記のような発話装置を皮膚表面電極33、34、37、38が設置されている部分が咽喉部分に当たるようにし、皮膚表面電極31、32、35、36、39、40、41、42を口の周辺に当たるように片手で持つ。そして、音節を発声（実際には音節に従って口を動かす）しながら、キーボードより発声音節の指示を行う。例えば、音節「く」を発声するように口を動かし、同時に、キーボードの「く」のキーを押下する。このような指示操作を所望の音節分繰り返す。

【0025】図5は、本発明の一実施例の発話装置の詳細な構成を示す。

【0026】発話装置1は、複数の対になった皮膚表面電極31～42、差動増幅器81～86、ローカットフィルタ91～96、ハイカットフィルタ101～106、FFT111～116、認識部3、音声合成部4、アンプ5及びスピーカ6より構成される。

【0027】発話装置1は、双極誘導するために各対になった皮膚表面電極31から42によって、ユーザの口周辺及び咽の筋内で生じる皮膚表面筋電信号を検出し、作動増幅器81～86で増幅され、ローカットフィルタ91～96により低周波を濾波し、ハイカットフィルタ101～106により各チャンネル毎に独立して高周波を濾波を行う。その後、FFT111～116において、時系列筋電信号をパワースペクトルに変換し、オクターブ解析したパワースペクトルの各バンド別成分に分解し、認識部3の神経回路網へ各バンド別成分が入力される。

【0028】図6は、本発明の一実施例の認識部と音声合成部の詳細な構成を示す。

【0029】同図において、認識部3は、入力層301、中間層302、出力層303より構成されるニューラルネットワークで構成され、○印で示されているものは、各ユニットを示し、○印と○印を接続している線は、結合荷重を含むリンクを表す。認識部3は、FFT111～116のいずれかのチャンネルから入力された各バンド別成分に基づいて、ユーザにより発声意図された音節を認識し、認識された音声のユニットが発火し、音声合成部4に音節信号が送られる。音声合成部4は、認識部3から出力された音節信号により音声の合成を行い、アンプ5に出力する。

【0030】アンプ5により合成された音声信号は増幅され、スピーカ6から音節信号を出力する。

【0031】以下に発話者が発話装置を用いて指示動作

(4)

特開平7 181888

5

(学習処理)を行う動作について説明する。

【0032】図7は、本発明の一実施例の教示動作を示すフローチャートである。

【0033】ステップ101)まず、認識部3のニューラルネットワークの各ユニットとリンクの結合荷重を乱数で初期化する。

【0034】ステップ102)発話者が口を動かして発声動作を行って顔面の筋肉が動くことにより、これを双極電極31~42の何れかの電極が筋電信号として検知し、取り込みを開始する。

【0035】ステップ103)このとき、発話者は、発声を意図する音節のキーボード2のキーを押下する。これにより時系列筋電信号に対して開かれた時間窓内(図8. a)において、ローカットフィルタ91~96のいずれかのフィルタが低周波成分を除去し、また、ハイカットフィルタ101~106のいずれかのフィルタが高周波成分を除去し、これらの周波数成分が除去された筋電信号がFFT111~116のいずれかに入力される。FFT111~116は、筋電信号をパワースペクトルに変換するFFT(高速フーリエ変換)処理を行い、認識部3に入力する。

【0036】ステップ104)一方、キーボード2で押下された意図された音節は(図8においては「あ」)、認識部3の出力層に教師信号として与えられる。認識部3への入力信号である教示音節信号70と教師信号であるキーボード2からの入力信号の誤差を図8に示す。図8において、いま発声音節が「あ」の場合には、認識部3の入力層301にFFTされたバンド別にパワースペクトル化された図9に示すような教示音節信号70が入力される。図9において、教示信号70(FFT111~116によりバンド別にFFTされ、パワースペクトル化した、筋電信号を0~1に正規化された数値列の教示)を認識部3の入力層301の各ユニットに入力する。また、キーボード2からの入力信号である教師信号60を認識部3の出力層303の各ユニットに入力する。例えば、認識部3の出力層303には、発声音節「あ」に対応して「1. 0. 0. . . . 0」のように、「あ」にのみ「1」、他は「0」の信号が与えられる。

【0037】認識部3のニューラルネットワークの学習は、例えば、誤差逆伝播法(文献:中野聖豊「ニューロンコンピュータ、技術評論社、p47、1989」)によって行われる。時間窓は、図8に示すようにオーバーラップしながら推移され、キーボード2が押下されている間は、常に認識部3のニューラルネットワークに教示音節信号が与えられ続けられる。

【0038】ステップ105)認識部3の学習処理のための教示は、全ての候補の音節が終了するまで繰り返す。上記のステップを繰り返す。

【0039】認識部3による上記の教示による学習が終

6

了した後は、キーボード2を発話装置1から外し、ユーザは発話装置1のみを用いて、任意の音節の発声動作(認識処理)を行う。即ち、発話装置1から音声合成で意図された発声音節が出力され、他者とのコミュニケーションを図ることができる。

【0040】次に、通常使用時(認識処理)の信号処理の手順を説明する。教示が既に終了している場合には、発話者が任意の発声を行うことにより、発話装置1は、発声された音声の認識を行う。

10 【0041】図10は、本発明の一実施例の発声動作の認識処理のフローチャートである。同図に示す認識処理では、既に教示(学習処理)が終了していることを前提とする。

【0042】また、認識処理時の筋電信号の処理を図11に示す。

【0043】ステップ201)まず、筋電信号の取り込みを開始する。

【0044】ステップ202)発話者が発話装置1から発声動作を行う。

20 【0045】ステップ203)皮膚表面電極31~42は、発話者の顔面の筋肉の動作に応じて、筋電信号を検出する。図11に示すように時系列に記される筋電信号sは、 $t=t1$ のとき、ローカットフィルタ91~96、ハイカットフィルタ101~106により濾波及びウィンドウ処理された後、FFT111~116に与えられ、認識部3に入力される。 $t=t1$ の筋電信号のバンド別パワースペクトルは同図に示すように認識部3に与えられ、当該筋電信号がどの音節に対応しているかを認識する。

30 【0046】ステップ204)認識部3により認識された信号は、 $t=t2$ に至るまでの間に音声合成部4から例えば、音声「あ」として出力される。

【0047】 $t=t2$ の時も同様に時系列筋電信号は、濾波及びウィンドウ処理された後、FFT処理され、同じ認識部3に $t=t2$ の時の入力として与えられ、認識処理の後、例えば、発声「あ」が継続されているときには、音声合成部4から「あ」として発声代行される。

【0048】ステップ205)以後、13、14...についても同様の処理が行われ、発声が意図された音節を音声合成部4から出力される。

【0049】本発明は、上記の実施例に限定されることなく、応用分野としては、無聲聴者の発声代行装置としても応用の他に、音声信号を認識する音声認識の補助手段として既存の音声認識手段と併用して、音声認識率を上げる応用や、静寂が必要で声を出せない環境において、発声はせずに、発声動作のみして口述の文書作成に応用する、高騒音環境下で既存音声認識では音声認識できない時に利用する等の種々応用が考えられる。

【0050】

50 【発明の効果】上述のように、本発明の発話装置を使用

(5)

特開平7-181888

7

8

する場合に、ユーザが予め、自分の筋電パターンと発声を意図する音節の関係を発話装置に学習させることにより、ユーザ個人のパターンに適応した認識ができ、学習能力のない認識装置を使用するのみ比較して、認識率が高いという利点がある。

【0051】また、ユーザ個人に適応しているので、ユーザが新たに発声動作を訓練する必要がなく、自然な発話動作の維持が期待できる。

【0052】さらに、複数の皮膚表面電極で検出した皮膚表面筋電信号を周波数ベクトルに分解し、バンド別成分を認識部で並列に処理するので、従来、困難であった筋電信号から音の識別も十分に可能である。

【図面の簡単な説明】

【図1】本発明の原理構成図である。

【図2】本発明の概要を説明するための図である。

【図3】本発明の一実施例のシステムを説明するための図である。

【図4】本発明の一実施例の発話装置の背面図である。

【図5】本発明の一実施例の発話装置の詳細な構成図である。

【図6】本発明の一実施例の認識部と音声合成部の詳細な構成図である。

【図7】本発明の一実施例の教示動作を示すフローチャートである。

【図8】本発明の一実施例の認識処理を説明するための図である。

【図9】本発明の一実施例の認識部に対する入力を示す図である。

\*【図10】本発明の一実施例の発声動作の認識処理のフローチャートである。

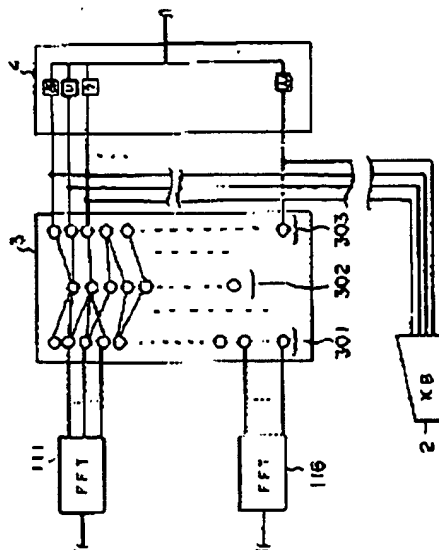
【図11】本発明の一実施例の認識処理時の信号状態を示す図である。

【符号の説明】

- 1 発話装置
- 2 キーボード
- 3 認識部
- 4 音声合成部
- 5 アンプ
- 6 スピーカ
- 10 信号検出手段
- 11 音節識別手段
- 12 学習手段
- 13 認識手段
- 14 音声合成手段
- 15 音声出力手段
- 31~42 皮膚表面電極
- 60 教師信号
- 70 筋電信号
- 81~86 差動増幅器
- 91~96 ローカットフィルタ
- 101~106 ハイカットフィルタ
- 111~116 FFT
- 301 入力層
- 302 中間層
- 303 出力層

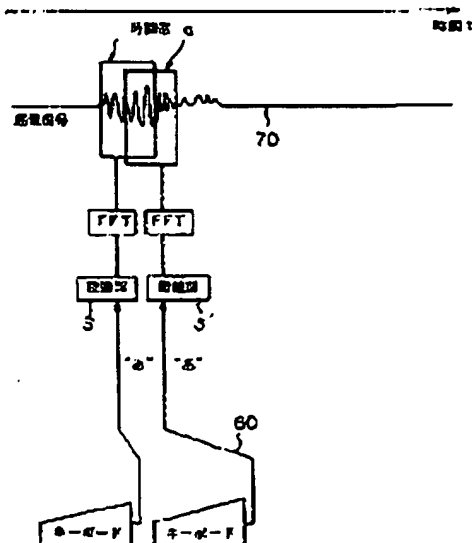
【図6】

本発明の一実施例の認識部と音声部の詳細な構成図



【図8】

本発明の一実施例の認識処理を説明するための図

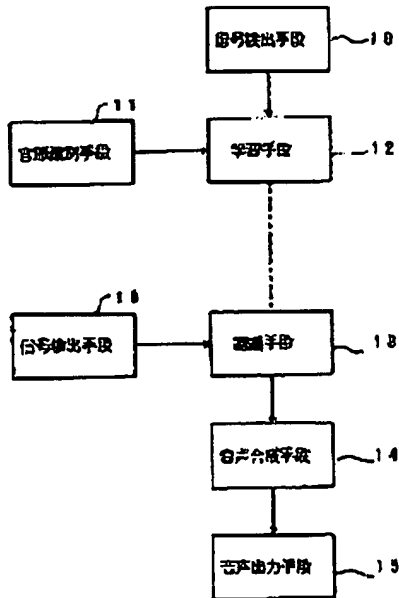


(6)

特開平7-181888

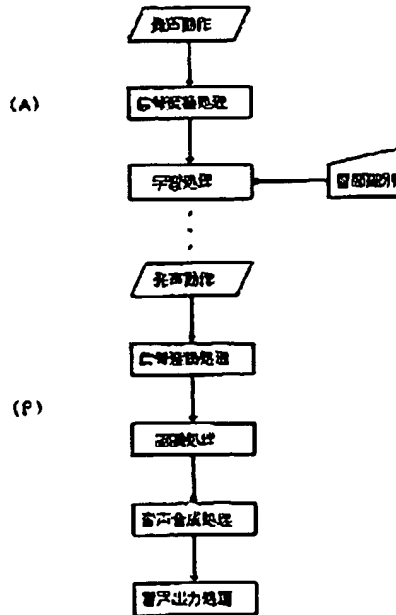
【図1】

本発明の処理構成図



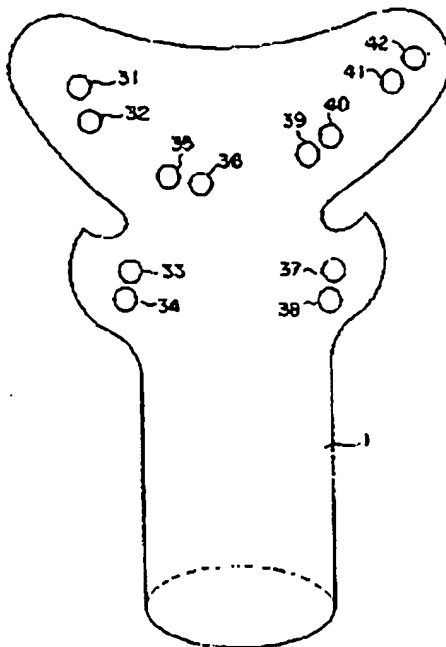
【図2】

本発明の概要を説明するための図



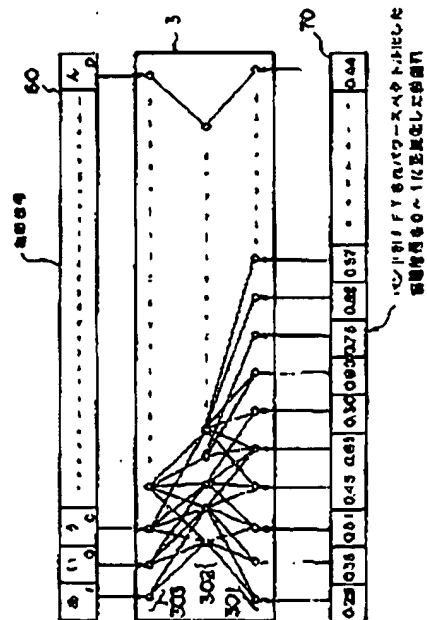
【図4】

本発明の一実施例の発光装置の正面図



【図9】

本発明の一実施例の発光装置に対する入力を示す図

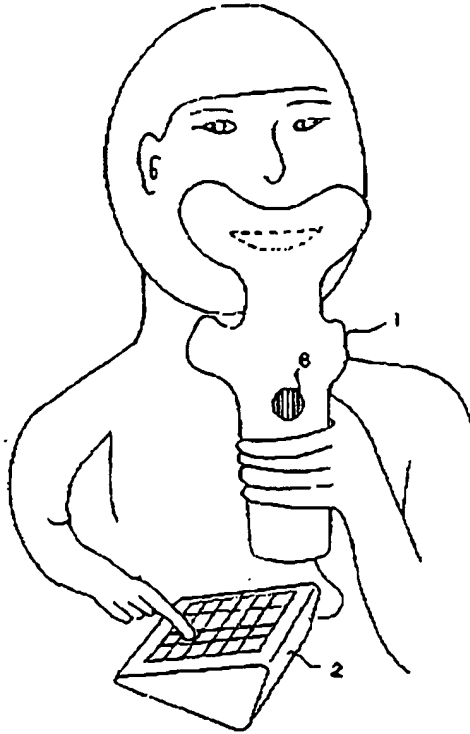


(7)

特開平7-181888

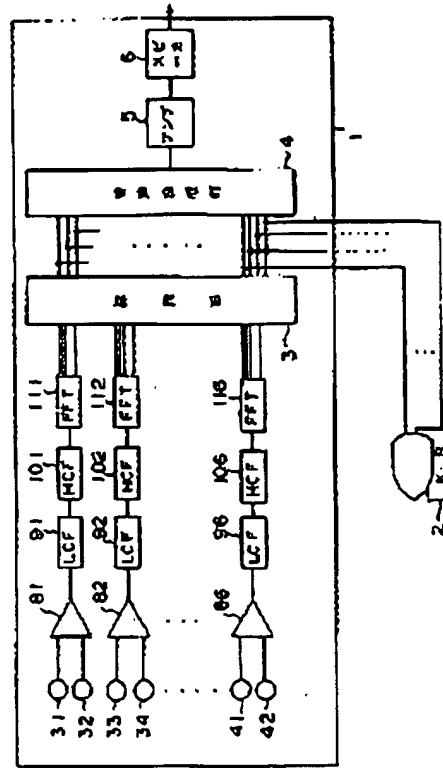
【図3】

本発明の一実施例のシステムを説明するための図



【図5】

本発明の一実施例の発振回路の構成図

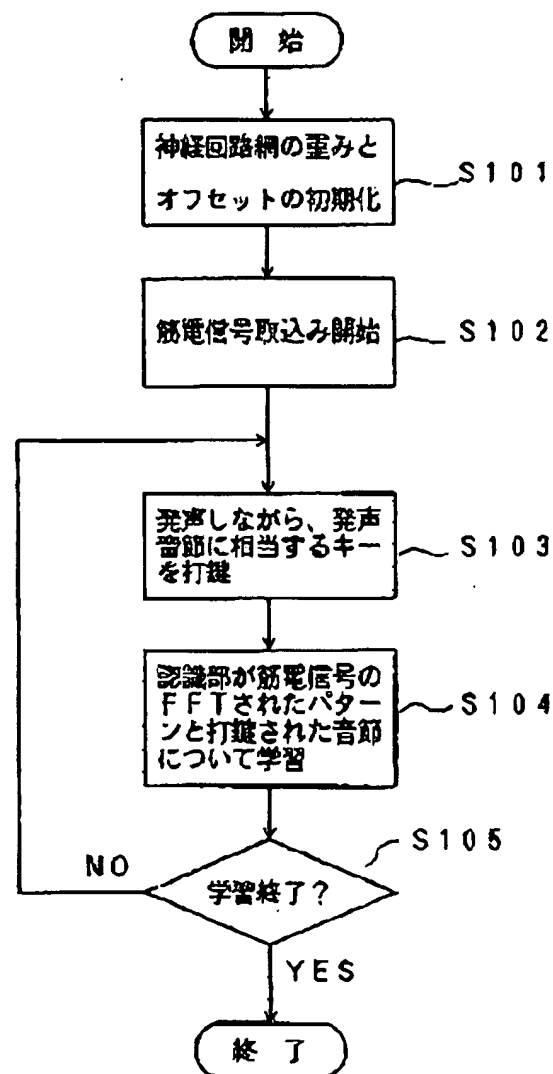


(8)

特開平7 181888

【図7】

本発明の一実施例の教示動作を示すフローチャート



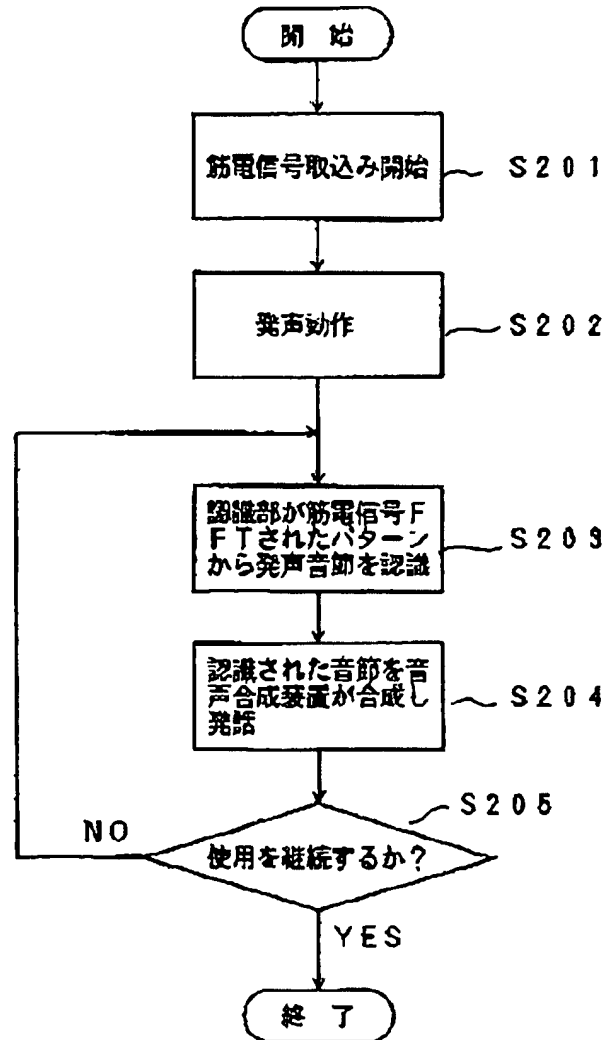


(7)

特開平7-181888

[図10]

本発明の一実施例の発声動作の認識処理フローチャート

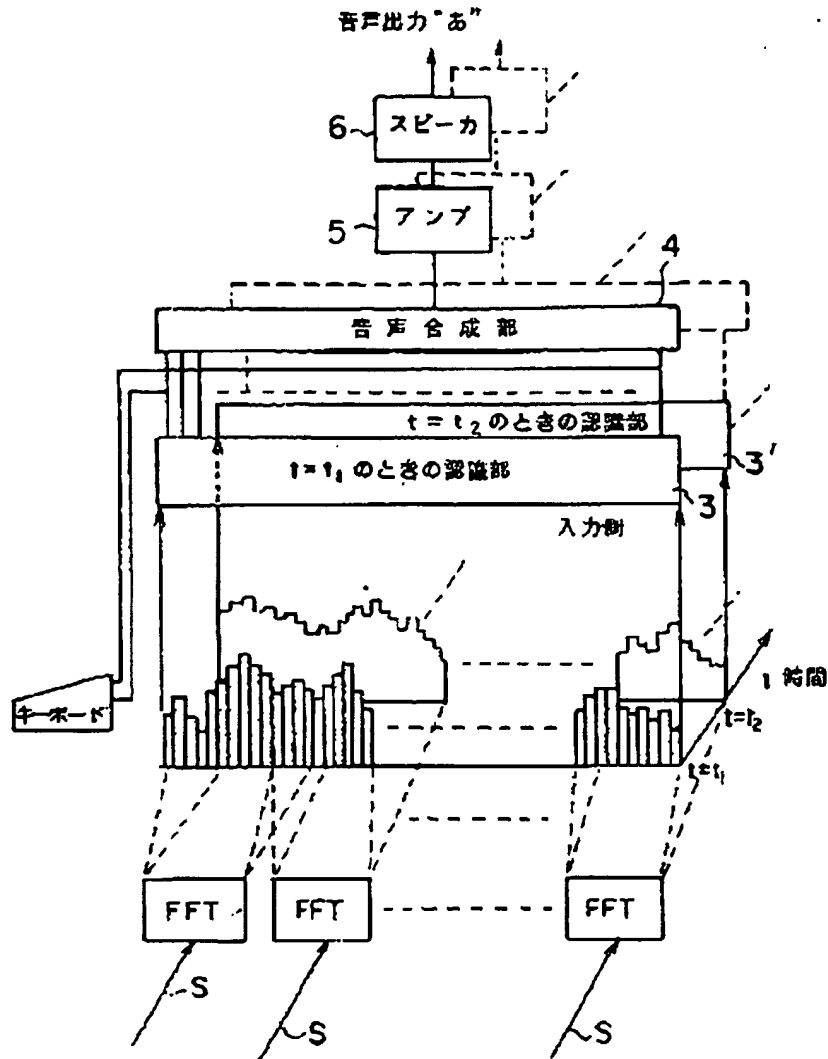


(1)

特開平 7-181883

【図 1】

本発明の一実施例の認識処理時の信号状態を示す図



(11) Publication No. JP H07-181888

(43) Date of Publication of Application 07.21.1995

---

**CLAIMS**

---

**[Claim(s)]**

[Claim 1] Phonation vicarious execution equipment characterized by providing the following. A signal-detection means to detect the signal generated by the movement of the muscles by phonation operation. A syllable discernment means to generate the recognition signal corresponding to desired syllable while performing this phonation operation. A study means to make into an input configuration the signal detected by this signal-detection means, and to learn by using as a teacher pattern the recognition signal corresponding to this syllable generated by this voice discernment means. A recognition means to make an input configuration the signal detected by this signal-detection means, and to recognize and output the phonation syllable signal corresponding to this input configuration, a speech synthesis means to compound this phonation sound signal outputted by this recognition means, and to change into voice, and a voice output means to output as voice this sound signal changed by this speech synthesis means.

[Claim 2] The aforementioned signal-detection means is phonation vicarious execution equipment including the amplification means for amplifying the myo-electric-signal signal detected from two or more skin surface electrodes and these skin surface electrodes for the movement of the skin detecting a signal, a filter means to remove the low-frequency component and high frequency component of the signal amplified by this amplification means, and a signal transformation means to change into a power spectrum the signal filtered by this filter means according to claim 1.

[Claim 3] The aforementioned study means and a recognition means are phonation vicarious execution equipment according to claim 1 which uses a neural network.

---

**DETAILED DESCRIPTION**

---

**[Detailed Description of the Invention]****[0001]**

[Industrial Application] As opposed to the non-pharynx person who this invention requires for phonation vicarious execution equipment, receives ablation of the pharynx with a certain illness especially, and cannot do phonation. The skin surface myo-electric-signal pattern detected from the skin surface electrode with which a gena and the mandible section were equipped when you wanted to utter syllable and it meant is recognized. The syllable which meant voice is recognized and it is related with the phonation vicarious execution equipment which changes to the artificial pharynx which synthesizes voice by recognizing the myo-electric-signal pattern which utters the recognized voice artificially by the voice synthesizer.

**[0002]**

[Description of the Prior Art] Although normal phonation becomes impossible to the non-pharynx person who lost the pharynx with a certain illness, the method of applying vibrator out of a throat and sending a sound source into the pharynx, and the method of inserting a pipe into a mouth and sending it directly are to give phonation of these men.

[0003] Moreover, the movement of a phonation person's lip is incorporated by the picture, and there is lip leading by the computer which carries out recognition processing.

[0004] Furthermore, there is research of processing the myo-electric-signal signal generated in connection with the movement of the muscles around a mouth, and discriminating a vowel kind. This Reference "Noboru Sugie et. al. and A speech Employing a Speech Synthesizer Vowel Discrimination from Perioral Muscles It is shown in Activities and Vowel Production, IEEE transactions on Biomedical Engineering, Vol.32, No.7, and pp485-490." After letting a band pass filter pass for a myo-electric-signal signal, the number of times of intersection of a threshold is counted, and it discriminates from five vowels (a, i, u, e, o).

[0005]

[Problem(s) to be Solved by the Invention] However, each of methods of sending in a sound source using the above-mentioned conventional vibrator, and methods using a pipe has unnatural sound, and since it becomes sound like a buzzer, other persons cannot recognize the sound easily as voice.

[0006] Moreover, lip leading by the computer is dependent on that the recognition rate by the image processing is low, and the recognition capacity of a computer.

[0007] Furthermore, although the method of discriminating a vowel with a myo-electric-signal signal can discriminate a vowel, it has about a consonant the problem of not being completely discriminable.

[0008] this invention was made in view of the above-mentioned point, solves the above-mentioned conventional trouble, and although it cannot utter normally by a certain reason, it aims at offering the phonation vicarious execution equipment which can output the synthesized speech based on the myoelectric potential of the mouth circumference to a sake.

[0009]

[Means for Solving the Problem] Drawing 1 is the principle block diagram of this invention.

[0010] A signal-detection means 10 to detect the signal generated by the movement of the muscles by phonation operation, A syllable discernment means 11 to generate the recognition signal corresponding to desired syllable while performing phonation operation, A study means 12 to make into an input configuration the signal detected by the signal-detection means 10, and to learn by using as a teacher pattern the recognition signal corresponding to the syllable generated by the voice discernment means 11, A recognition means 13 to make an input configuration the signal detected by the signal-detection means 10, and to recognize and output the phonation syllable signal corresponding to an input configuration, It has a \*\*\*\*\* means 14 to compound the phonation sound signal outputted by the recognition means 13, and to change into voice, and a voice output means 15 to output as voice the sound signal changed by the speech synthesis means 14.

[0011] Moreover, the signal-detection means 10 of this invention includes the amplification means for amplifying the myo-electric-signal signal detected from two or more skin surface

electrodes and skin surface electrodes for the movement of the skin detecting a signal, a filter means to remove the low-frequency component and high frequency component of the signal amplified by the amplification means, and a signal transformation means to change into a power spectrum the signal filtered by the filter means.

[0012] Moreover, a neural network is used for the study means 12 and the recognition means 13 of this invention.

[0013]

[Function] this invention introduces into the recognition section the neural network who is a processor with a high recognition rate strong against study nature and noise, inputs the signal acquired from phonation operation used as study data by this neural network, and the syllable correspondence signal corresponding to phonation operation as teacher data, and performs study processing. Change into frequency spectrum after study the myo-electric-signal signal detected by phonation operation of a user from two or more skin surface electrodes by FFT (fast Fourier transform), and the changed spectrum is recognized. The meant voice syllable is recognized, it can synthesize voice with the myo-electric-signal signal by phonation operation of a user, and the phonation vicarious execution equipment which carries out a voice output can be offered, and when a non-pharynx person performs phonation operation by everyday life, the synthesized speech near natural voice can be outputted.

[0014]

[Example] Hereafter, the example of this invention is explained with a drawing.

[0015] The outline of this invention is explained first.

[0016] Drawing 2 is drawing for explaining the outline of this invention.

[0017] First, as shown in this drawing (A), a user performs utterance operation by moving a mouth for desired syllable using utterance equipment, and the signal according to the operation is detected. In order to use this signal as study data, signal transformation processing is performed, and it inputs into a neural network. Furthermore, a recognition signal is generated with utterance operation using the device which can create the signal corresponding to syllable, such as a keyboard, and let this be teacher data of study processing. A neural network learns with the study data and teacher data which were inputted.

[0018] Study processing is ended, and in order for the user who performed study processing to perform transform processing so that the signal which performed utterance operation from utterance equipment and was acquired by this operation can be inputted into a neural network, and to perform recognition processing, it inputs into a neural network. Since the neural network is learning previously, he recognizes the inputted signal and outputs a recognition result. It synthesizes voice from the signal which it is as a result of recognition, and outputs as voice.

[0019] In this example, although utterance operation is performed using utterance equipment, a user does not utter voice, but this detects the signal uttered by moving the surrounding muscles of a mouth, and performs study processing and recognition processing based on this signal.

[0020] Drawing 3 is drawing for explaining the system of one example of this invention.

[0021] As shown in this drawing, utterance equipment 1 is connected with a keyboard 2,

and a user performs the instruction input about the voice which had utterance equipment 1 single hand and was uttered from the keyboard 2 by the hand of another side. Utterance equipment is the case which consisted of plastics of non-insulation etc., as shown in this drawing, and it is constituted by the configuration which consists of a curve so that near an utterance person's mouth and the portion of a throat may be contacted.

[0022] Drawing 4 is drawing which looked at the utterance equipment of one example of this invention from the rear face.

[0023] This drawing is drawing at the time of seeing the utterance equipment 1 shown in drawing 2 from a rear face, and the skin surface electrodes 31-42 are arranged in the rear face of utterance equipment 1, each electrodes 31-42 are pairs in order to carry out bipolar guidance like electrodes 31 and 32 and electrodes 33 and 34, and a myo-electric-signal signal is acquired by equipping the front face of an utterance person's face.

[0024] It is made for the portion in which the skin surface electrodes 33, 34, 37, and 38 are installed in the above utterance equipments to be equivalent to a throat portion, and an utterance person has the skin surface electrodes 31, 32, 35, 36, 39, 40, 41, and 42 single hand so that it may hit around a mouth. And phonation syllable is taught from a keyboard, uttering syllable (according to syllable, a mouth is moved in fact). For example, a mouth is moved so that syllable "\*\*\*" may be uttered, and the depression of the key of "\*\*\*" of a keyboard is carried out simultaneously. It repeats by the syllable of a request of such instruction operation.

[0025] Drawing 5 shows the detailed composition of the utterance equipment of one example of this invention.

[0026] Utterance equipment 1 consists of the skin surface electrodes 31-42 which became two or more pairs, differential amplifier 81-86, low cut filters 91-96, high frequency cut filters 101-106, FFT 111-116, the recognition section 3, the speech synthesis section 4, amplifier 5, and a loudspeaker 6.

[0027] Utterance equipment 1 detects the skin surface myo-electric-signal signal produced with a user's mouth circumference and the muscles of a throat by the skin surface electrodes 31-42 which became each set in order to carry out the bipolar lead, is amplified with the operation amplifier 81-86, filters low frequency by low cut filters 91-96, and performs Rf filtering independently for every channel by high frequency cut filters 101-106. Then, in FFT 111-116, a time series myo-electric-signal signal is changed into a power spectrum, it decomposes into each component according to band of the power spectrum which carried out octave analysis, and each component according to band is inputted into the neuron network of the recognition section 3.

[0028] Drawing 6 shows the detailed composition of the recognition section of one example of this invention, and the speech synthesis section.

[0029] In this drawing, what the recognition section 3 consists of neural networks who consist of an input layer 301, an interlayer 302, and an output layer 303, and is shown by O mark shows a unit respectively, and the line which has connected O mark and O mark expresses a link including a joint load. The recognition section 3 recognizes the syllable in which the phonation intention was done by the user based on each component according to band inputted from one channel of FFT 111-116, the unit of the recognized voice ignites, and a syllable signal is sent to the speech synthesis section 4. The speech synthesis section

4 compounds voice with the syllable signal outputted from the recognition section 3, and outputs it to amplifier 5.

[0030] The sound signal compounded with amplifier 5 is amplified, and outputs a syllable signal from a loudspeaker 6.

[0031] Operation whose utterance person uses utterance equipment for below, and performs instruction operation (study processing) to it is explained.

[0032] Drawing 7 is a flow chart which shows instruction operation of one example of this invention.

[0033] Step 101 The joint load of each unit of the neural network of the recognition section 3 and a link is first initialized by random numbers.

[0034] When a step 102 utterance person moves a mouth and the muscles of a line and the face move utterance operation, which group of bipolar electrodes 31-42 detects this as a myo-electric-signal signal, and incorporation is started.

[0035] Step 103 At this time, an utterance person does the depression of the key of the keyboard 2 of the syllable which means utterance. The myo-electric-signal signal with which one filter of the low cut filters 91-96 removed the low-frequency component in the time window opened to the time series myo-electric-signal signal by this ( drawing 8 , a), and one filter of the high frequency cut filters 101-106 removed the high frequency component, and these frequency components were removed is inputted into either of FET 111-116. FET 111-116 performs FFT (fast Fourier transform) processing which changes a myo-electric-signal signal into a power spectrum, and inputs it into the recognition section 3.

[0036] Step 104 On the other hand, the meant syllable which was pushed by the keyboard 2 is given to the output layer of "\*\*\*\*") and the recognition section 3 as a teacher signal in ( drawing 8 . The outline of the input signal from the keyboard 2 which is the instruction syllable signal 70 and teacher signal which are an input signal to the recognition section 3 is shown in drawing 8 . In drawing 8 , when phonation syllable is "\*\*\*\*" now, the instruction syllable signal 70 as shown in drawing 9 power-spectrum-ized according to the band by which FFT was carried out to the input layer 301 of the recognition section 3 is inputted. In drawing 9 , the instruction signal 70 (numeric value of the numerical train in which FFT was carried out according to the band by FFT 111-116 and which was normalized by 0-1 in the power-spectrum-ized myo-electric-signal signal) is inputted into each unit of the input layer 301 of the recognition section 3. Moreover, the teacher signal 60 which is an input signal from a keyboard 2 is inputted into each unit of the output layer 303 of the recognition section 3. the output layer 303 of the recognition section 3 -- phonation syllable "\*\*\*\*" -- corresponding -- "1, 0 and 0, and ... as for "1" and others, the signal of "0" is given to "\*\*\*\*" like .0"

[0037] Study of the neural network of the recognition section 3 is performed for example, by the error reverse spreading method (47 reference : Nakano \*\*\*\*\* a neuron computer, technical Ilyoronsha, p 1989). While it changes while a time window overlaps, as shown in drawing 8 , and the keyboard 2 is pushed, it always continues being given in an instruction syllable signal at the neural network of the recognition section 3.

[0038] The instruction for study processing of the step 105 recognition section 3 repeats the above-mentioned step repeatedly until the syllable of all kinds is completed.

[0039] After study by the above-mentioned instruction by the recognition section 3 is completed, a keyboard 2 is removed from utterance equipment 1, and a user performs phonation operation (recognition processing) of arbitrary syllable only using utterance equipment 1. That is, the phonation syllable meant by speech synthesis from utterance equipment 1 is outputted, and communication with the others can be aimed at.

[0040] Next, the procedure of signal processing at the time of use (recognition processing) is usually explained. When instruction is already completed and an utterance person performs arbitrary phonation, utterance equipment 1 recognizes uttered voice.

[0041] Drawing 10 is the flow chart of recognition processing of phonation operation of one example of this invention. It is premised on instruction (study processing) already being completed in the recognition processing shown in this drawing.

[0042] Moreover, processing of the myo-electric-signal signal at the time of recognition processing is shown in drawing 11.

[0043] Step 201 The incorporation of a myo-electric-signal signal is started first.

[0044] A step 202 utterance person performs phonation operation from utterance equipment 1.

[0045] The step 203 skin surface electrodes 31-42 detect a myo-electric-signal signal according to operation of the muscles an utterance person's face. At the time of  $t=t1$ , the myo-electric-signal signal is allotted to time series as shown in drawing 11 is given to FFT 111-116 by low cut filters 91-96 and high frequency cut filters 101-106, filtering and after windowing is carried out, and it is inputted into the recognition section 3. The power spectrum according to band of the myo-electric-signal signal of  $t=t1$  is given to the recognition section 3 as shown in this drawing, and it recognizes to which syllable the myo-electric-signal signal concerned corresponds.

[0046] By the time the signal recognized by the step 204 recognition section 3 results in  $t=t2$ , it will be outputted as voice "\*\*\*" from the speech synthesis section 4.

[0047] Similarly, when FFT processing is carried out filtering and after windowing is carried out, a time series myo-electric-signal signal is given to the same recognition section 3 as an input at the time of  $t=t2$  and utterance "\*\*\*" is continued after recognition processing, the utterance vicarious execution also of the time of  $t=t2$  is carried out as "\*\*\*" from the speech synthesis section 4.

[0048] After step 205, processing with the same said of  $t3$  and  $t4$  -- is performed, and the syllable by which utterance was meant is outputted from the speech synthesis section 4.

[0049] this invention, without being limited to the above-mentioned example as an applicable field in the environment where it is required and voice cannot be uttered the application which uses together with the existing speech recognition means as auxiliary means of speech recognition which recognize the sound signal other than the application also as a non-pharynx person's utterance vicarious execution equipment, and gathers the rate of speech recognition, and calm \*\* -- As for utterance, using, when speech recognition cannot be carried out in the existing speech recognition under the high noise environment which carries out only utterance operation and it applies to oral statement document preparation, without carrying out etc. can consider application variously.

[0050]

[Effect of the Invention] As mentioned above, when using the utterance equipment of this



invention, and a user makes utterance equipment learn beforehand the relation of the syllable which means its own myo-electric-signal pattern and utterance first, recognition which was adapted for a user individual's pattern can be performed, it accepts and compares and there is an advantage which uses recognition equipment without machine learning that a recognition rate is high.

[0051] Moreover, since the user individual is fitted, a user does not newly need to train utterance operation and maintenance of natural utterance operation can be expected.

[0052] Furthermore, since the skin surface myo-electric-signal detected by two or more skin surface electrodes is decomposed into frequency PEKUTORU and the component according to band is processed in parallel in the recognition section, discernment of the difficult myo-electric-signal signal to a consonant is also fully possible conventionally.

---

## DESCRIPTION OF DRAWINGS

---

### [Brief Description of the Drawings]

[Drawing 1] It is the principle block diagram of this invention.

[Drawing 2] It is drawing for explaining the outline of this invention.

[Drawing 3] It is drawing for explaining the system of one example of this invention.

[Drawing 4] It is the rear view of the utterance equipment of one example of this invention.

[Drawing 5] It is the detailed block diagram of the utterance equipment of one example of this invention.

[Drawing 6] It is the detailed block diagram of the recognition section of one example of this invention, and the speech synthesis section.

[Drawing 7] It is the flow chart which shows instruction operation of one example of this invention.

[Drawing 8] It is drawing for explaining recognition processing of one example of this invention.

[Drawing 9] It is drawing showing the input to the recognition section of one example of this invention.

[Drawing 10] It is the flow chart of recognition processing of utterance operation of one example of this invention.

[Drawing 11] It is drawing showing the signal state at the time of recognition processing of one example of this invention.

### [Description of Notations]

1 Utterance Equipment

2 Keyboard

3 Recognition Section

4 Speech Synthesis Section

5 Amplifier

6 Loudspeaker

10 Signal-Detection Means  
11 Syllable Discernment Means  
12 Study Means  
13 Recognition Means  
14 Speech Synthesis Means  
15 Voice Output Means  
31-42 Skin surface electrode  
60 Teacher Signal  
70 Myo-Electric-Signal Signal  
81-86 Differential amplifier  
91-96 Low cut filter  
101-106 High frequency cut filter  
111~116 FET  
301 Input Layer  
302 Interlayer  
303 Output Layer